# Student privacy self-management: implications for learning analytics

## Conference or Workshop Item

# Student privacy self-management: implications for learning analytics

Paul Prinsloo
University of South Africa
3-15, Club 1, P O Box 392, Unisa
0003, South Africa
+27 12 4334719
prinsp@unisa.ac.za

Sharon Slade
Open University
Foxcombe Hall, Boars Hill
Oxford, UK
+44 1865 327000
sharon.slade@open.ac.uk

## ABSTRACT
Optimizing the harvesting and analysis of student data promises to clear the fog surrounding the key drivers of student success and retention, and provide potential for improved student success. At the same time, concerns are increasingly voiced around the extent to which individuals are routinely and progressively tracked as they engage online. The Internet, the very thing that promised to open up possibilities and to break down communication barriers, now threatens to narrow it again through the panopticon of mass surveillance.

Within higher education, our assumptions and understanding of issues surrounding student attitudes to privacy are influenced both by the apparent ease with which the public appear to share the detail of their lives and our paternalistic institutional cultures. As such, it can be easy to allow our enthusiasm for the possibilities offered by learning analytics to outweigh consideration of issues of privacy.

This paper explores issues around consent and the seemingly simple choice to allow students to opt-in or opt-out of having their data tracked. We consider how 3 providers of massive open online courses (MOOCs) inform users of how their data is used, and discuss how higher education institutions can work toward an approach which engages and more fully informs students of the implications of learning analytics on their personal data.

## Categories and Subject Descriptors
K.3.1 [Computers and Education]: Computer Uses in Education - *Distance learning, K.7.4 [The Computing Profession]: Professional Ethics – Codes of ethics*

## General Terms
Management, Documentation, Security, Legal Aspects.

## Keywords
Learning analytics, ethics, informed consent, opt out, opting out


LAK '15, March 16 - 20, 2015, Poughkeepsie, NY, USA

http://dx.doi.org/10.1145/2723576.2723585

## 1. INTRODUCTION
The growing use and enthusiasm for learning analytics mirrors both the rapid development of technologies and the increased understanding of how student data may be effectively applied to improve learning and student success [9, 10, 16, 54, 55].

However, the emergence of learning analytics as an approach which underpins teaching and learning strategy has in turn raised awareness amongst students and other stakeholders regarding both the uses to which student data may be put and the approaches employed within learning analytics [10, 56, 57].The issue of student privacy self-management is furthermore embedded in definitions of privacy [24, 30, 39, 58, 67] and concerns that "discussions of privacy merely scratch the surface" and we need "a better understanding of the problems; we must learn how they developed, how they are connected, what precisely they threaten, and how they can be solved" [58]. Our definitions, as well as our legal and regulatory frameworks often struggle to keep up with technological developments and changing societal norms [64].

Amidst increasing concerns around ethical issues in the harvesting, analysis and use of student data resides the question of whether students should have the opportunity to 'opt out' of institutional surveillance. What are the implications of learning analytics and student surveillance on the social contracts that higher education institutions (HEIs) have with their stakeholders, and, in particular, their students [29, 31]? How does allowing students to opt out of the harvesting, analysis and use of their personal data impact on the fiduciary duty of HEIs [14, 34] given their responsibility to ensure appropriate support and guidance to students in their learning journeys? Do students understand the benefits and/or costs of opting in or out at the moment they make a choice? [e.g., 22].

As the uses to which student data are put continue to evolve, it is becoming clear that existing policies may be failing to keep pace with those changes. In previous research, Prinsloo and Slade [51] investigated the extent to which current policies at two mega distance education institutions address the harvesting, analysis and use of student data. The study found that, while informed student consent was an established principle when data is used for *research* purposes [42], student consent was not required in the day-to-day harvesting and use of student data. The collection, analysis and use of student data was also scattered between a number of policies and role players.

Furthermore, until recently most of the data harvested and analysed has been aggregated and issues of anonymity and privacy have not been problematic. Learning analytics has changed this. The potential of learning analytics lies in harvesting personalised

data to intervene, where appropriate and possible, at an *individual* level. This raises a number of ethical and practical concerns [22, 42, 57].

Some HEIs are responding by creating specific policy around their use of learning analytics, for example, the Open University in the UK released a new policy on the ethical uses of learning analytics to students and staff in October 2014 [50]. Although such policies do much to address transparency and boundary concerns, there remain further issues around the specific area of informed consent and the possibilities of students opting out of having their data used.

As far as we could establish, there is little or no published research or theorising in the context of higher education whether students should have the option to opt-out of having their learning data harvested, analysed and used, often in combination with other sets of data. Though Buchanan [6] and Slade and Prinsloo [56, 57] point to the possibility of increasing transparency and trust by providing an option to opt-out; there is no published research on the feasibility or implications of such an option. In one consultation reported by Slade and Prinsloo [57], student representatives clearly indicated that they would expect to have an option to opt-in or opt-out.

Though the notion of opting out is yet to be seriously considered in higher education; the discourse is well-established in fields such as legal and surveillance studies [1, 2, 24, 39, 40, 47, 60, 61] Stoddart [62] suggests that it is possible to approach the issue of 'privacy self-management' either from a "rights-based or discursive-disclosive" approach. The first approach sets out procedural guarantees establishing rules and access to satisfaction if these rules are breached. Stoddart [62] and others indicate that this approach appears increasingly ineffective. The "discursive-disclosive" approach situates surveillance in the context of what is being done, by whom, and for what purpose and then to investigate alternative approaches to satisfy the need that resulted in surveillance.

It is unlikely that increasing ethical concerns and issues may be addressed so simply. The impact of technologies and the "elaborate lattice of information networking" [60], suggests that legal frameworks may never be enough to completely safeguard users against possible abuse and misuse of their data [52].

In this paper we will situate the question of allowing students to opt-in or out against an overview of discourses regarding privacy self-management and the sharing of personal data. An analysis of the Terms and Conditions (TOCs) of three providers of massive open online courses (MOOCs) provides a sense of the accessibility of these TOCs, the nature and scope of data harvesting, the use and sharing of data, and issues regarding user access to their own data. Based on the literature review and analysis, we then formulate a number of theses for consideration in learning analytics.

# 2. SOME EXISTING APPROACHES TO PRIVACY SELF-MANAGEMENT

It falls outside the scope of this paper to discuss the vast literature (historical and current) addressing the notion of privacy [24, 30, 39, 53, 58, 67]. Privacy self-management and consent is therefore a fluid and changing notion worth exploring. It is also important that legal and regulatory frameworks differ between geopolitical and institutional contexts.

Solove [59] states that "Consent is an under theorized concept that is crucial for privacy and many other areas of law. Consent performs an enormous amount of work. Activities that would otherwise be illegitimate are made legitimate by consent". Most of the current models informing our assumptions about consent and privacy self-management are primarily based on a paternalistic approach. There are two main arguments against an uncritical adoption of such an approach:

(1) The correctness of our choices regarding privacy and data use is not always clear [569], and is dependent on context [49]. Despite assumptions to the contrary, there are individuals who actively welcome targeted marketing. "These people should not be dismissed as uninformed or foolish, as it is far from clear that the costs to these people outweigh the benefits" [59]. Individuals may share personal data willingly in one context for a specific purpose, but feel that their privacy was disregarded if the same information was shared with a wider audience or for another purpose. An example of the latter is the sharing of often intensely personal medical information for the greater benefit of others on PatientsLikeMe. There is also ample evidence of the ways in which changing social norms regarding the scope and nature of personal data shared can result in changes in how we see and/or participate in surveillance, self-surveillance or 'the quantified self' phenomenon, or sousveillance [32, 38, 41].

(2) We cannot disregard the fact that "the collection, use, and disclosure of personal data - even without consent - can lead to great benefits for individuals and society" [59]. There is, however, also ample evidence where the indiscriminate or compulsory collection of personal data results in the disenfranchising of already vulnerable populations and individuals [26, 28]. Solove therefore states that "Paternalism is much easier to justify when the consequences [of not collecting data] are clearly bad" [59].

A discursive-disclosive approach tells "both us and others what we did not previously know about our situations, the conditions under which we have been living and working and how we might be being exploited as, for example, in the extensive cross-referencing of personal data…" [62]. We should therefore examine how we are encouraged, or coerced into sharing personal data. The outcome of this approach is not to declare surveillance either good or bad, but to understand "how it is dangerous and where the dynamic of resistance-compliance by all parties might be more effectively practiced" [62].

## 2.1 Explicit opt in or opt out

Many of the current models informing privacy self-management involve an explicit opt in/opt out choice (see for example [45]). Bellman et al [3] explore how the framing of questions and response defaults can impact on the opt-in or opt-out choices made by consumers. When provided with a choice to opt-in to receive marketing information, a majority of respondents wanted to make that choice on a case-by-case basis. Phrasing options as opt-*out* instead of opting in also changed the privacy preferences. The research found that marketers would predetermine users' preferences by setting the default option as opting in when users don't indicate a specific preference. The font size and the technical detail and complexity of the privacy policy documents also had a huge impact on users' engagement and choices. It is therefore crucial to note their finding that "Opting-in does not equal opting-out, and answers are influenced by the default option" [3]. Based on their research, Bellman et al [3] suggest

that everything possible should be done to encourage users to indicate a *conscious* choice. Without definite answers from users, no data collection should take place.

While Solove initially supported the opt-in option, he questions the effectiveness and ethical foundations of opting in on the basis that "many organizations will have the sophistication and motivation to find ways to generate high opt-in rates" [59]. Often, as in the case with Apple's iTunes, opting in is a prerequisite to basic use. "Despite the knowledge benefit of opt-in, *there is a cost.* In opt-in regimes, people affirmatively indicate their consent to data collection and sharing. With this clearer and more legitimate consent, companies might feel more entitled to use and disclose data more widely. In contrast, with opt-out, the consent procured is less legitimate than with opt-in regimes" [59; emphasis added].

## 2.2 Increasing understanding and agency

Despite the challenges, gaps, and unresolved issues of privacy self-management, Solove [59] suggests that it should not be abandoned – "Providing people with notice, access, and the ability to control their data is key to facilitating some autonomy in a world where decisions are increasingly being made about them with the use of personal data, automated processes, and clandestine rationales, and where people have minimal abilities to do anything about such decisions". Indeed the main beneficiary of increasing a broader understanding of the importance of privacy self-management is not necessarily the individual user, but rather companies. The need to address the practical implications of being transparent regarding the collection, analysis and use of personal data can force companies and corporations to address processes and clarify their own thinking.

Zuckerman [71] and others [28, 64] point to an increase in user and consumer activism regarding the collection and use of personal data. The growth in "privacy enhancing technologies" (PETs) such as TrackMeNot, Tor, DuckDuckGo and Ghostery allow individuals to obfuscate or prevent the collection of search histories and other data. While privacy is "not complete control of our personal information, nor is it perfect secrecy" (Nissenbaum quoted by Zuckerman, [71]). Zuckerman [71] therefore proposes increased civic efficacy which involves advocacy to change laws and provide greater access to privacy protection and access to redress; as well as seeking change through markets, shaping social norms and through code.

# 3. PROBLEMATISING AGENCY IN PRIVACY SELF-MANAGEMENT

The question of whether to allow students to opt-out of having their learning data harvested and analysed seems out of place in the current social climate of "digital promiscuity" [48], where "sharing is caring" and "privacy is theft" [18] and given the prevalence of student indiscretions on social media [69]. Authors such as Kitchen [32] map the different actors in the field of use of personal data and there is clear indication that sousveillance and the sharing of personal information is increasing. Murphy [48] points to the irony that, in an age where we give up more of our personal information for free, we are increasingly worried about privacy. Despite, or rather, amidst the seemingly pervasive digital promiscuity, there are signs of "people beginning to exercise a bit more reserve online or otherwise engaging in subversive tactics to thwart data miners" [48]. (Also see [46, 63, 65].)

## 3.1 The notion of informed consent

Solove [59] states that "Privacy self-management envisions an informed and rational person who makes appropriate decisions about whether to consent to various forms of collection, use, and disclosure of personal data". This vision of individuals' ability to comprehend the issues at stake and to make rational decisions is perhaps far removed from reality.

Solove [59] suggests that people "often lack enough expertise to adequately assess the consequences of agreeing to certain present uses or disclosures of their data. People routinely turn over their data for very small benefits" (also see [65]). This points to "a clear disconnect between people's expressed high value of privacy and their behavior, which indicates a very low value of privacy" [59].

Despite a general belief that individuals base their decisions on rational processes, we underestimate the "bounded rationality" people have as they negotiate meaning and compromise priorities within complex situations [59]. Trepte and Reinecke [65] emphasise that users make decisions regarding the sharing of data based on a cost-benefit analysis, shaped by their understanding of the costs and benefits, based on the amount and correctness of information they have to their disposal. Wang [66] therefore claims that users often trade privacy for convenience. There may be different layers of costs and benefits – and as Nissenbaum [49] indicates, context plays a crucial role in making decisions to opt-in or opt-out. These contexts may also be dynamic and the use of personal data in one context or in a particular time may not necessarily be appropriate in another context or time.

Users' choices and their understanding of the scope and effectiveness of their privacy self-management are developed in context rather than in the abstract and "People are also more willing to share personal data when they feel in control, regardless of whether that control is real or illusory" [59].

It is clear from a review of literature that there are concerns that most TOCs and policies on privacy are not read by most users, or if they are, are not well understood in terms of the regulations, costs and benefits. Users may lack the necessary background knowledge to make an informed choice. A further complicating factor is the different norms governing different context impact on the simplicity and context of the binary to opt-in or out at a particular moment in time [4, 7, 11, 25, 37, 49, 61].

## 3.2 Problems with informed consent

There are a number of impediments impacting on individuals' ability to consent to or opt-out of the various forms of collection, use, and disclosure of personal data. These include practical problems with the notion and practice of informed consent such as the problem of scale, re-identification and the problem of aggregation, quality, accountability and purpose of data, limitations on use, openness and user agency.

### *3.2.1 The problem of scale*

It is almost impossible for an individual to comprehend not only the scope of data collected, analysed and used, but the implications of the different layers of collection, analysis and use, with specific reference to third parties and other entities. "Even if every entity provided people with an easy and clear way to manage their privacy, there are simply too many entities that

collect, use, and disclose people's data for the rational person to handle" [59].

It therefore becomes almost impossible for users to engage with each site's terms and conditions in order to make informed choices. No matter how clear or transparent policies and terms and conditions are, it has become impossible due to the sheer scale of the collection, analysis and use.

### *3.2.2 Re-identification and the problem of aggregation*

The relationship between privacy, identity and anonymity is more nuanced than often presumed, and "what appears at first blush to be a zero-sum game is in fact a set of interdependent relationships" [30]. The notion of opting out seems to imply that anonymity will ensure privacy – but increasingly this is not the case. Central to the "murky conceptual waters" of distinguishing (and regulating these distinctions) between the public and the private [44] is the need to recognise privacy as a social and political construct [68]. Not only does the notion of privacy have different connotations throughout history [53], it involves "multiple meanings over time and across cultures, contexts, kinds of persons and social categories" [44]. Privacy is embedded in existing socio-economic power relations and is increasingly a form of currency in "informational capitalism" [12].

Increasing complexity and a growing awareness of that complexity can often trigger a drive toward either further regulation or to the adoption of a paternalistic approach. Solove [59] suggests that "consent to collection, use, and disclosure of personal data is often not meaningful, but the most apparent solution - paternalistic measures - even more directly denies people the freedom to make consensual choices about their data". A paternalistic approach may have been an appropriate solution in contexts where the use of data "had little benefit or were primarily detrimental to the individual or society". Generally, the use of data is complex as the same set of data may have both benefits and costs for the individual or groups from which it was harvested. Different individuals within a group may then come to different conclusions regarding whether the costs of having that data collected outweighs the benefits.

Privacy self-management has its origins in the Fair Information Practice Principles (FIPPs). Though Marx [43] states that the FIPPs are no longer adequate, for the purpose of this paper it is useful to briefly recap them. The principles include: "(1) transparency of record systems of personal data, (2) the right to notice about such record systems, (3) the right to prevent personal data from being used for new purposes without consent, (4) the right to correct or amend one's records, and (5) responsibilities on the holders of data to prevent its misuse". Solove [59] suggests that the FIPPs fail to clarify "what data may be collected or how it may be used. Instead, most forms of data collection, use, and disclosure are permissible under the FIPPs if individuals have the ability to self-manage their privacy — that is, if they are notified and provide consent". Addressing the shortcomings of the FIPPs, Marx [43] proposed 29 questions to guide institutions and individuals through the ethical issues in surveillance. The questions covered aspects such as the means of collection, the data collection context, and uses of the collected data (see [56] for a discussion of Marx's proposal and a framework for ethical learning analytics).

Solove [59] suggests that rational decisions made regarding distinct datasets may struggle to include how that data might be aggregated in future. Individuals may share innocuous data in particular contexts, but it is the aggregation of these data points that reveals a picture that cannot be anticipated or foreseen at the moment of sharing.

While the aggregation of data can also be beneficial to individuals, "it is virtually impossible for a person to make meaningful judgments about the costs and benefits of revealing certain data" [59]. There are too many unknowns with regard to how data may be combined at a certain future point in time, when the original context in which the data was captured is no longer known. A cost-benefit analysis that includes a range of possible harms and benefits is not only almost impossible to do, but is also dependent on context [49].

A presumed advantage of aggregation is the assumption of anonymity. However, while a single data point may not be personally identifiable, the aggregation of a number of data points may actually result in the identification of the individual involved. "As data gets aggregated, information that is not identifiable can become identified" [59].

### *3.2.3 Quality, accountability and purpose of data*

Big Data refers not only to increasingly high volumes, but also to data's velocity. It incorporates both structured and unstructured data, and may be exhaustive in scope, striving to capture both entire populations of systems and fine-grained resolution, and have an inherent ability to be combined and related to other and different data-sets [32]. In this emerging world then, it becomes easier for the original purpose for the collection of any dataset to be lost as "new uses for data are discovered over time" [8]. As data becomes further removed from its original purpose and context of collection, there may be impacts on its quality and fit within its new context. With the increased brokering of data, it also becomes almost impossible to trace and enforce accountability. We therefore witness an increasing context-collapse in the use of data.

### *3.2.4 Limitations on use*

Cate and Mayer-Schönberger [8] state that, although there is agreement that the "Use limitation principle" adopted in 1980 is no longer workable, it is not clear what should replace it. According to this principle, personal data should not be used, disclosed, nor made available, other than for the original purpose of the collection, except with the explicit consent of the data subject or by the authority of law.

### *3.2.5 Openness*

Should data be collected without opportunities to get consent, Cate and Mayer-Schönberger [8] suggest that "information about the data subjects' legal rights, and ways to exercise them, be made available with information about data processing activities".

### *3.2.6 User agency*

Although there is general agreement with regard to individuals rights to know and to have access to data, there are concerns regarding to the "significant burdens for both data processors and data subjects in a world of Big Data" [8]. There are basically three concerns. Firstly, with the amount of data collected as well as the different role players (known and unknown), "access to

such data could be prohibitively expensive and seemingly of little value if the data are not being used for any significant purpose", for example, as part of broad-based research rather than a requirement to determine a specific benefit [8]. The second concern relates to the burden on individuals to establish the different parties with access to their data, and the possibility that data may have been de-identified for use. Thirdly, by focusing on individual agency and participation, the burden on data stewards may be excessive without really addressing the need for fundamental fairness in data. See, for example Marx [43].

# 4. RESEARCH DESIGN AND METHODOLOGY

Content analysis as methodology is an established methodology in quantitative and qualitative research designs [5, 15, 19, 27]. Hsieh and Shannon [27] describe content analysis as "a family of analytical approaches ranging from impressionistic, intuitive, interpretative analysis to systematic, strict textual analyses". Qualitative content analysis "is defined as a research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns". Elo and Kyngäs [19] state that the aim of content analysis is "to attain a condensed and broad description of the phenomenon, and the outcome of the analysis is concepts or categories describing the phenomenon" resulting in a "model, conceptual system, conceptual map or categories".

In following Elo and Kyngäs [19], we developed a categorisation matrix based on constructs formulated from the literature review. Validity, reliability and trustworthiness in deductive content analysis, are furthermore addressed by transparency regarding the process including the selection of analytical constructs from the literature review, coding, member checking of the codes, constructs and analyses [15, 19, 70].

The credibility of this study was based on frameworks provided by Elo and Kyngäs [19] and Lincoln and Guba [35, 36]. A thorough literature review resulted in a categorization matrix and referential adequacy and member checking and peer debriefing. Transferability was ensured through thick descriptions and reference to the source documents. Process notes and an electronic paper trail provide an audit trail.

Three providers of massive open online courses (MOOCs) were selected for this study, namely Coursera [13], edX [17], and FutureLearn [21]. The Terms and Conditions of the three providers exported into an MSWord file, and analysed through coding and the formulation of themes.

The analytical constructs used in the analysis included the following:

a. Length of Terms and Conditions
b. Types of data collected
c. Methods of data collection
d. Conditions for sharing the data collected
e. Uses of data
f. User access to, responsibility and control of data
g. Institutional duty of care

# 5. ANALYSIS AND DISCUSSION OF FINDINGS

## 5.1 Length of Terms and Conditions (TOC)

It is clear from the literature review, e.g., [3, 59] that most users of online services do not read the TOCs due to the amount of reading as well as the sometimes technical nature of the documents. An analysis of the length of the TOC (including privacy policies and honor codes) of the three providers show that the documents were between 5,965 words or 13 pages (Coursera) and 8,565 words or 22 pages (FutureLearn). The documents cover a wide variety of topics, in different orders, levels of detail and uses of technical terms. It falls outside the scope of this paper to compare the content of these TOCs, but it is interesting to note that the number of headings in these TOCs range between 30 headings and subheadings in the Coursera TOC to 41 in the edX TOC. Another interesting aspect of a typographic analysis of these texts shows that the amount of text in bold and capitals (excluding headings) range from 460 (edX), 784 (Coursera) and none for FutureLearn. Research [23] has shown that typefaces impact not only on the readability of the text but also on the emotions created in readers with bold and capitalized text being seen as more serious.

## 5.2 Types of data collected

Privacy policies for most organisations involved in education discuss their use of both *personally identifiable* data (for example, contact details and other demographic information) and *non-personal* data (for example, webpages visited, etc.).

### *5.2.1 Personally identifiable data*

Only edX provides a definition of personal information while Coursera contrasts personal information with non-personal information as "information that cannot be used to identity you." All three providers require a range of personal information from users in order to register for courses, such as email addresses, contact details, gender, and date of birth. edX also require a photograph to verify students' identities for accreditation purposes. All three providers make it very clear that they collect the personally identifiable information that users directly provide. Interestingly, FutureLearn also includes the following "*we may receive* information about you from third parties (such as credit reference agencies) who are legally entitled to disclose that information" (Privacy Policy; emphasis added). Both Coursera and edX flag that their tracking of non-personal information may yield further information that could be also used to identify personal data, such as location or IP addresses. In contrast, FutureLearn's policy explicitly states that it will "collect data relating to forum posts which "cannot identify you".

### *5.2.2 Non-personal information*

Both Coursera and edX mention that they aggregate non-personal information, for example, sites visited, and hyperlinks 'clicked.' FutureLearn does not refer at all to aggregation of data, but states that "We may also collect data relating to your visits to the Website that cannot identify you but records your use of our Website".

## 5.3 Methods of data collection

All three providers mention the use of cookies to collect personal and non-personal information. Coursera, however, adds that

"From time to time we may also use additional typical methods of collecting data". No further information is provided with regard to the type of data or methods. Though all three providers indicate that users *can* disable cookies, they each warn that this may impact on the functionality of the service provided. FutureLearn has a separate Cookie Policy which it also contains an overview of the different types of cookies and even provides a list of cookies used in the collection of data.

## 5.4 Conditions for sharing the data collected

Coursera does not specify any conditions under which collected personalised data will be used and/or shared. With regard to sharing non-personalised information it states "We may also use it for other business purposes". Coursera goes on to state that it may use personally identifiable information collected on the Forums and it "may publish this information via extensions of our Platform that use third-party services, like mobile applications".

Both edX and FutureLearn have more information regarding the conditions under which collected data will be shared. edX will share collected data with affiliated universities on condition that the information is treated in a confidential manner and protected. Only personal information required to "fulfill the purpose stated at the time of collection" will be shared. FutureLearn states that personalised information will be shared with the consent of the individual in question, or to prospective sellers or buyers of assets, or should it be compelled under legal obligations to do so. These last two conditions are also covered by Coursera and edX. FutureLearn is the only provider who mentions obtaining the consent of the relevant individual.

## 5.5 Uses of data

Coursera does not have a specific section dealing with the uses of data, but lists potential uses under the types of data collected. All three providers included information regarding the collection of data with the purpose to provide, administer, evaluate and improve the offerings and services. The providers also acknowledged that they collect data to understand (FutureLearn), facilitate (Coursera) and individualise the learning/browsing experience (edX, FutureLearn).

Coursera and edX mention specifically the collection of data to authenticate the user's identity and track both individual and aggregate attendance, progress and completion.

edX is the only provider to mention the collection of data to monitor and detect violations, to publish, for purposes made clear at the point of collection. edX and FutureLearn mention explicitly that data may be collected for research purposes while Coursera state that data may also be collected for business purposes. Somehow related to this is FutureLearn's acknowledgement that collected data may be used to sell sponsorship.

Coursera states that identifiable information generated by students, such as forum postings, may be both published and later shared with others (through reuse). edX goes a little further to state that any forum posts are fully owned by edX in perpetuity and may be later exploited in whole or in part. Forum posts are analysed to obtain information about student performance and patterns of learning, but it is not clear whether this information is de-personalised or may be linked directly to student identity and other personal data. Not only that, but any other user of the edX site (both current and in the future) is granted license to access and use student posts (which may contain the student username) for their own personal purposes. Perhaps because of this, edX encourages students to adopt usernames that keep their identities concealed, but, requires at the same time, that students must commit not to misrepresent their identities. FutureLearn, on the other hand, insists that students use real names as identifiers and encourages users to openly share (with them and with other students) details of their location, gender and education history to "help other learners get to know you and help us to tailor the service to suit you".

## 5.6 User access to, responsibility and control of data

### *5.6.1 Opting out is not an option*

Though all three providers are very transparent with regard to the fact that they collect, use and share data; none of the providers provide users the option to opt out of the collection of data, whether personalised data or in aggregated form. All three providers do provide users with the option to opt out of allowing cookies, but with the warning that this may impact negatively on the quality of the service provided.

### *5.6.2 The duty of care*

All three providers also make it very clear that, outside of the required personal information needed, they will use whatever personal information users provide to them. Users are also warned that once shared, the provider cannot guarantee the security and safety of the implications of the sharing of information. Coursera, for example, states "Please note that *we do not guarantee* the security of Personally Identifiable Information, and there is some risk that an unauthorized third party may find a way to circumvent our security systems or that transmission of your information over the Internet will be intercepted".

### *5.6.3 User responsibility for correctness of data*

All three providers make it clear that users have the responsibility to ensure that the required data provided is correct and current. FutureLearn, for example, states: "You can edit your personal details via your profile page whenever you wish. We maintain a procedure in order to help you confirm that your personal information remains correct and up-to-date or choose whether or not you wish to receive material from us or some of our partners". (See section 5.5 regarding impersonation and false identities). It warns that "You acknowledge that if any information provided by you in relation to your Learner Account is untrue, inaccurate, not current or incomplete, we reserve the right to suspend or terminate your access to and use of the Website and your enrolment in the Online Content and Courses"

### *5.6.4 User concerns regarding privacy and data use*

All three providers provide users an opportunity to raise concerns or question policy by sending an email, For example, edX states: "If you have privacy concerns, have disclosed data you would prefer to keep private, or would like to access the information we maintain about you, please contact us at privacy@edx.org".

Only FutureLearn provides users with the possibility to access the data it holds for a small fee, stating "You have the right to contact us in order to find out what information we hold about you … or to access, cancel or correct any information that we hold about you."

# 6. IMPLICATIONS FOR LEARNING ANALYTICS

## 6.1 The duty of reciprocal care

From the literature review, analysis and findings it is clear that the duty of care is shared between providing institutions and users – albeit embedded in an asymmetrical power relationship. In the TOCs of the three providers analysed, any suggestion of a balance of power would be nonsensical. The power to harvest, analyse and exploit data lies completely with the provider. It is tempting to state that the only real option available to users to have more control over their data is to not use the provider at all. However, it is also clear that this is not a realistic, nor an appropriate response.

What has become clear within the scope of duty and care is that both duty and care have reciprocal elements. Having said this, it is important to also note that, due to the heavy imbalance in the power relationship, the main responsibility lies with the provider to ensure transparency, security and reasonable care (see section 6.2 for further discussion).

The social contract and fiduciary duty of HEIs and other educational providers therefore necessitate that providers

- Make their TOCs as accessible and understandable as possible. There is a vast difference between reading the TOCs of Coursera and edX compared with FutureLearn – the latter being more accessible, albeit longer.
- Make it clear what data is collected, for what purposes, and with whom data may be shared (and under what conditions).
- Provide users with access to information and data held about them, and to verify or correct conclusions drawn, where necessary, as well as provide context, if appropriate.
- Provide users with access to a neutral ombudsperson who can ensure that concerns and issues raised are addressed in a consistent and just manner. Though the three providers in this study provide users with an email address, this is not enough. Complaints and concerns should be addressed to a neutral third person or agency.
- Ensure that users are provided with opportunities to verify and update personal information.

## 6.2 The contextual integrity of privacy and data

One of the major concerns in the discourses in surveillance and privacy studies is the issue of contextual integrity of privacy and personal data (see, for example [49]). Data and information that are collected and/or shared in one context lose contextual integrity when shared or used out of context.

## 6.3 Student agency and privacy self-management

Due to the fact that the power-relationship between students and their higher education institution is, per se, asymmetrical, it is crucial to embrace the proposal by Slade and Prinsloo [57] to see learning analytics as *moral* practice. Should the social contract between students and institution [29, 31] be seriously considered, learning analytics as moral practice provides form and function to the fiduciary [14, 34] and duty of care [42] that higher education has towards students. The social contract and fiduciary duty of care provides a crucial basis for thinking critically about the range of student control over what data will be analysed, for what purposes, and how students will have access to verify, correct or supply additional information [34]. When students are seen as active collaborators and agents in the harvesting, analysis and use of their data [57], the potential for thinking outside of the binary of opting in or out becomes a real possibility.

Despite concerns regarding the digital promiscuity [48] of students and their apparent lack of care with regard to uncritical sharing of information, HEIs have a moral obligation to not only make students aware of the implications, but also to provide a platform for empowering students with civic agency regarding their data.

From the above, it is clear that informed consent is not a simple matter, and that most users don't understand the full implications of opting in or out from allowing their data to be used and analysed. It is furthermore clear that student surveillance in higher education should not be thought of in terms of just good or evil [53], but as a necessary and crucial tool within the context of the social contract and duty of care.

HEIs should therefore not accept as default a non-response from users regarding the collection of data as equal to opting in. Bellman et al [3] suggest that "no data collection or use should occur until a definite answer has been received from the consumer". The costs and benefits of accepting the TOC should be clear to users at the outset [59].

## 6.4 Future direction and reflection

### *6.4.1 Rethinking Consent and Employing Nudges*

The "murky conceptual waters" [44] of distinguishing between public and private, and the fact that the notion of valid consent can vary depending on the area and jurisdiction of law considered, does not make easier the task of reconsidering the role of privacy self-management. It is not always clear to the user what granting consent to having one's data collected, analysed and used actually means. Considering the asymmetrical power relation between the service consumer or user and the provider, does the user really have an option? In line with the thinking of Trepte and Reinecke [65] that we should not underestimate the reciprocal aspect of data sharing and use, and issues of cost and benefit, how do we safeguard against misuse? Solove [59] therefore suggests that "consent is far more nuanced, and privacy law needs a new approach that accounts for the nuances without getting too complex to be workable".

One possible solution might be to employ an approach of nudging, or presumed consent, much discussed in the fields of health and energy policy. Nudging, which draw on approaches developed within behavioural economics, works to steer individual decision making so as to make individuals better off without breaching their free choice. It may be argued that nudge policies are based on principles of soft paternalism, in that they make it easier for people in act in ways that support both the policy owner's and the broader public's best interests. However, it is thought that a nudge approach may provide an acceptable middle ground between paternalism and privacy self-management.

### *6.4.2 Developing Partial Privacy Self-Management*

Accepting that users have the right to choose but may also not have the time nor necessarily the knowledge to micromanage every aspect of personal data use, a further potential solution

might be to embrace partial self-management. Solove [59] proposes that individuals might be able to "manage their privacy globally for all entities rather than one at a time". This, however, raises a different issue namely "to find a uniform set of privacy options that makes sense for all entities, and the consequences of data collection, use, or disclosure might differ depending upon which entities are involved" [59].

### *6.4.3 Adjusting Privacy's Timing and Focus*

The current emphasis is that individuals choose how they want their data to be used at the moment of using a particular service or network – and often, at that stage, they either don't care or simply don't have the information needed to make an informed decision. "Therefore, the focus should be more on downstream uses rather than on the time of the initial collection of data" [59]. The benefits in opting in or out may not be apparent at the time the data is collected. "New ideas for combining data, new discoveries in data aggregation and analysis, and new techniques and technologies of data analysis might change the benefits side of the equation. They might change the costs side as well" [59].

Solove [59] suggests that users may be provided different options such as outright restrictions, partial consent depending on the scope, context and timing, and permission to harvest and use data with an option to later revoke consent or change the scope of consent depending on the context or circumstances.

### *6.4.4 Moving Toward Substance over Neutrality*

Solove [59] further proposes that the law "should develop and codify basic privacy norms. Such codification need not be overly paternalistic — it can be in a form like the Uniform Commercial Code (UCC), where certain default rules can be waived" [59]. This will result in "More substantive rules about data collection, use, and disclosure could consist of hard boundaries that block particularly troublesome practices as well as softer default rules that can be bargained around" [59].

Certainly, within the European Union [20], the 1995 Data Protection Directive (95/46/EC), is being updated to reflect the ways in which technological progress and globalisation have profoundly changed the way data is collected, accessed and used. The draft regulation requires that consent to the processing of personal data be given explicitly; and a right for data subjects to be forgotten, including the right to obtain erasure of personal data available publicly online. The new regulation is expected to be adopted in 2014, with implementation two years later, in 2016. (Also take note of [45]).

## 7. CONCLUSIONS

Within higher education, the enthusiasm that surrounds the emergence of learning analytics and its ability to offer differentiated support can overwhelm the detail of its implementation, particularly with regard to issues of consent. HEIs, on the whole, have adopted a paternalistic approach, often electing either to not inform students of the ways in which their data is being used to track progress or to offer partial insight with no opportunity to opt out. However, it is clear that the way forward cannot simply be to introduce a choice between opt-in or opt-out.

HEIs must engage more proactively with students, to inform and more directly involve them in the ways in which both individual and aggregated data is being used. It should not be blithely assumed that HEIs can, or indeed will, be able to work for the benefit of individual students, and, indeed, who decides what is in the best interests of individuals? Only by increasing the transparency around learning analytics activities will HEIs gain the trust and fuller co-operation of students. As Solove [59] states "The way forward involves (1) developing a coherent approach to consent, one that accounts for the social science discoveries about how people make decisions about personal data; (2) recognizing that people can engage in privacy self management only selectively; (3) adjusting privacy law's timing to focus on downstream uses; and (4) developing more substantive privacy rules. These are enormous challenges, but they must be tackled."

## 8. REFERENCES


[1] Ball, K., Haggerty, K.D. and Lyon, D. 2012. In *Routledge handbook of surveillance studies*. Routledge, Abingdon, UK.

[2] Barnard-Wills, D. 2012. In *Surveillance and identity. Discourse, subjectivity and the state.* Ashgate Publishing Ltd, Farnham, UK.

[3] Bellman, S., Johnson, E.J. and Lohse, G.L. 2001. On site: to opt-in or opt-out?: it depends on the question. Communications of the ACM, 44, 2, (2001), 25-27. Retrieved from http://dl.acm.org/citation.cfm?id=359241

[4] Bennett, C.J. 2001. Cookies, web bugs, webcams and cue cats: patterns of surveillance on the world wide web. *Ethics and Information Technology*, 3, (2001), 197-210.

[5] Bos, W. and Tarnai, C. 1999. Content analysis in empirical social research, *International Journal of Educational Research*, 31, (1999), 659-671.

[6] Buchanan, E.A. 2011. Internet research ethics: past, present and future. In *The handbook of Internet studies,* M.Consalvo and C.Ess, Eds. John Wiley, Oxford, UK, 83-108.

[7] Brandimarte, L., Acquisti, A. and Loewenstein, G. 2013. Misplaced confidences: privacy and the control paradox. *Social Psychological and Personality Science*, 4, (2013), 340-347. DOI: 10.1177/1948550612455931

[8] Cate, F.H. and Mayer-Schőnberger, V. 2013. Tomorrow's privacy. Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), (2013), 67-73.

[9] Clow, D. 2013a. An overview of learning analytics, *Teaching in Higher Education,* 18, 6, (2013), 683-695. DOI: 10.1080/13562517.2013.827653.

[10] Clow, D. 2013b, November 13. Looking harder at Course Signals, *Clow Blog.* Retrieved from http://dougclow.org/2013/11/13/looking-harder-at-course-signals/

[11] Cockcroft, S. 2006. Information privacy: Culture, legislation and user attitudes. *Australasian Journal of Information Systems*, 14, 1, (2006), 55-68.

[12] Coll, S. 2014. Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. *Information, Communication & Society,* 17, 10, (2014), 1250-1263. DOI: 10.1080/1369118X.2014.918636

[13] Coursera. 2014. Terms and conditions. Retrieved from https://authentication.coursera.org/auth/auth/normal/tos.php

[14] DeAngelis, W. 2014. Academic deans, codes of ethics, and … fiduciary duties? *Journal for Academic Ethics*, 12, (2014), 209-225. DOI: 10.1007/s10805-014-9212-4

[15] Downe-Wamboldt, B. 1992. Content analysis: method, applications, and issues, *Health Care for Women International,* 13, 3, (1992), 313-321. DOI: 10.1080/07399339209516006

[16] Drachsler, H. and Greller, W. 2014. The pulse of learning analytics understandings and expectations from the stakeholders. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, (2014), 120-129. Retrieved from http://dl.acm.org/citation.cfm?id=2330634

[17] edX. 2014. Terms of service (including Privacy policy). Retrieved from https://www.edx.org/edx-terms-service (Terms of service)

[18] Eggers, D. 2013. The circle. Penguin, London, UK.

[19] Elo, S. and Kyngäs, H. 2007. The qualitative content analysis process, Journal of Advanced Nursing, 62, (2007), 107-115. DOI: 10.1111/j.1365-2648.2007.04569.x.

[20] EU data protection framework (draft). Retrieved from http://www.parliament.uk/business/publications/research/briefing-papers/SN06669/the-draft-eu-data-protection-framework

[21] FutureLearn. 2014. Terms and conditions (including Privacy and Cookie Policy). Retrieved from https://about.futurelearn.com/terms/

[22] Gavison, R. 1980. Privacy and the limits of law. *Yale Law Journal,* 89, 421-471.

[23] Gump, J.E. 2001. The readability of typefaces and the subsequent mood or emotion created in the reader, *Journal of Education for Business,* 76, 5, (2001), 270-273. DOI: 10.1080/08832320109599647

[24] Haggerty, K.D. and Ericson, R.V. (Ed.). 2006. *The new politics of surveillance and visibility*. University of Toronto Press, Toronto, Canada.

[25] Haynes, A.W. 2006. Online privacy policies: contracting away control over personal information? *Penn State Law Review*, 111, 3, (2006), 587-624.

[26] Henman, P. 2004. Targeted!: Population segmentation, electronic surveillance and governing the unemployed in Australia. *International Sociology*, 19, (2004), 173-191. DOI: 10.1177/0268580904042899

[27] Hsieh, H-F. and Shannon, S.E. 2005. Three approaches to qualitative content analysis, *Qualitative Health Research,* 15, 9, (2005), 1277-1288. DOI: 10.1177/1049732305276687

[28] Irwin, J. 2014, October 7. Grooming students for a lifetime of surveillance. [Web log post]. Retrieved from http://modelviewculture.com/pieces/grooming-students-for-a-lifetime-of-surveillance

[29] Jongbloed, B., Enders, J. and Salerno, C. 2008. Higher education and its communities: interconnections, interdependencies and a research agenda. *Higher Education*, 56, (2008), 303-324.

[30] Kerr, I. and Barrigar, J. 2012. Privacy, identity and anonymity. In *Routledge Handbook of Surveillance Studies,* K.Ball, K.D. Haggerty and D.Lyon, Eds., Routledge, Abingdon, UK, 386-394. .

[31] Kharouf, H., Sekhon, H. and Roy, S.K. 2014. The components of trustworthiness for higher education: a transnational perspective. *Studies in Higher Education.* DOI: 10.1080/03075079.2014.881352

[32] Kitchen, R. 2013. Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography,* 3, (2013), 262-267.

[33] Kruse, A. and Pongsajapan, R. 2012. Student-centered learning analytics. CNDLS Thought Papers. Retrieved from https://cndls.georgetown.edu/m/documents/thoughtpaper-krusepongsajapan.pdf

[34] Lee, B.A. 2014. Student-faculty academic conflicts: emerging legal theories and judicial review. *Mississippi Law Journal*, 83, (2014), 837-951.

[35] Lincoln, Y.S. and Guba, E.G. 1985. *Naturalistic inquiry.* London, UK: SAGE Publications.

[36] Lincoln, Y. S. and Guba, E. G. 1990. Judging the quality of case study reports. International *Journal of Qualitative Studies in Education,* 3, 1, (1990), 53-59. DOI: 10.1080/0951839900030105

[37] Liu, C., Marchewka, J.T., Lu, J. and Yu, C-S. 2004. Beyond concern: a privacy-trust-behavioral intention model of electronic commerce. *Information & Management,* 42, (2004), 127-142. DOI: 10.1016/j.im.2004.01.002

[38] Lupton, D. 2012, November 4. The quantified self-movement: some sociological perspectives. [Web log post]. Retrieved from http://simplysociology.wordpress.com/2012/11/04/the-quantitative-self-movement-some-sociological-perspectives/

[39] Lyon, D. (Ed.). 2006. *Theorising surveillance. The panopticon and beyond.* Cullumpton, UK: Willan Publishing.

[40] Lyon, D. 2007. *Surveillance studies. An overview. Cambridge*, UK: Polity Press.

[41] Mann, S. 2012, November 2. Eye am a camera: surveillance and sousveillance in the glass age. *Time*. Retrieved from http://techland.time.com/2012/11/02/eye-am-a-camera-surveillance-and-sousveillance-in-the-glassage/

[42] Marshall, S. 2014. Exploring the ethical implications of MOOCs. *Distance Education,* 35, 2, (2014), 250-262. DOI: 10.1080/01587919.2014.917706

[43] Marx, G.T. 1998. Ethics for the new surveillance. *The Information Society: An International Journal*, 14, 3, (1998), 171-185. DOI: 10.1080/019722498128809

[44] Marx, G.T. 2001. Murky conceptual waters: the public and the private. *Ethics and Information Technology*, 3, (2001), 157-169.

[45] McAleese, M et al 2014. Report to the European Commission on new modes of learning and teaching in Higher Education, October 2014 retrieved from http://ec.europa.eu/education/library/reports/modernisation-universities_en.pdf

[46] Mohamed, N. and Ahmad, I.H. 2012. Information privacy concerns, antecedents and privacy measure use in social

networking sites: evidence from Malaysia. *Computers in Human Behavior*, 28, (2012), 2366-2375.

[47] Mont, M.C., Pearson, S., Bramhall, P. 2003. Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services. Retrieved from http://pdf.aminer.org/000/151/916/towards_accountable_management_of_identity_and_privacy_sticky_policies_and.pdf

[48] Murphy, K. 2014, October 4. We want privacy, but can't stop sharing. *The New York Times*. [Web log post]. Retrieved from http://www.nytimes.com/2014/10/05/sunday-review/we-want-privacy-but-cant-stop-sharing.html?partner=rss&emc=rss&smid=tw-nytopinion

[49] Nissenbaum, H. 2004. Privacy as contextual integrity. *Washington Law Review*, 79, (2004), 119-158.

[50] Open University. 2014. Retrieved from http://www.open.ac.uk/students/charter/essential-documents/ethical-use-student-data-learning-analytics-policy.

[51] Prinsloo, P. and Slade, S. 2013. An evaluation of policy frameworks for addressing ethical considerations in learning analytics. Paper presented at Learning Analytics and Knowledge Leuven, Belgium, 8-12 April, 2013. Retrieved from http://dl.acm.org/citation.cfm?id=2460344

[52] Raab, C.D. 2012. Regulating surveillance. The importance of principles. In *Routledge Handbook of Surveillance Studies,* K.Ball, K.D. Haggerty and D.Lyon, Eds., Routledge, Abingdon, UK, 377-385.

[53] Sewell, G. and Barker, J.R. 2001. Neither good, nor bad, but dangerous: surveillance as an ethical paradox. *Ethics and Information Technology,* 3, (2001), 183-196.

[54] Siemens, G. and Long, P. 2011. Penetrating the fog: Analytics in learning and education, *EDUCAUSEreview*, [online].September/October, Retrieved from http://www.elmhurst.edu/~richs/EC/OnlineMaterials/SPS102/Teaching%20and%20Learning/Penetrating%20the%20Fog.pdf

[55] Siemens, G. 2013. Learning analytics: The emergence of a discipline, *American Behavioural Scientist,* 57, 10, (2013), 1380-1400.

[56] Slade, S. and Prinsloo, P. 2013. Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*. DOI: 10.1177/0002764213479366

[57] Slade, P. and Prinsloo, P. 2014. Student perspectives on the use of their data: between intrusion, surveillance and care. Paper presented at 8th EDEN Research Workshop, 27-28 October, Oxford, UK.

[58] Solove, D.J. 2004. The digital person. Technology and privacy in the information age. New York, NY: New York University

[59] Solove, D.J. 2013. Introduction: Privacy self-management and the consent dilemma. Harvard Law Review 1880 (2013); GWU Legal Studies Research Paper No. 2012-141; GWU Law School Public Law Research Paper No. 2012-141. Available at SSRN: http://ssrn.com/abstract=2171018

[60] Solove, D.J. 2004 *The digital person.* New York University Press, New York, NY.

[61] Sovern, J. 1999. Opting in, opting out, or no options at all: The fight for control of personal information. *Washington Law Review,* 74, (1999), 1033-1130. Retrieved from http://netcaucus.org/books/privacy2001/pdf/Sovern.pdf

[62] Stoddart, E. 2012. A surveillance of care. Evaluating surveillance ethically. In *Routledge Handbook of Surveillance Studies,* K.Ball, K.D. Haggerty and D.Lyon, Eds., Routledge, Abingdon, UK, 369-376.

[63] Taddei, S. and Contena, B. 2013. Privacy, trust and control: which relationships with online self-disclosure? *Computers in Human Behaviour*, 29, (2013), 821-826.

[64] Tene, O, and Polonetsky, J. 2012. Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property,* 239, (2012), 1-36. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149364

[65] Trepte, S. and Reinecke, L. 2013. The reciprocal effects of social network site use and the disposition for self-disclosure: a longitudinal study. *Computers in Human Behavior,* 29, (2013), 1102-1112.

[66] Wang, R. 2013, June 10. Beware trading privacy for convenience. [Web log post]. Harvard Business Review. Retrieved from http://blogs.hbr.org/2013/06/beware-trading-privacy-for-con/

[67] Warren, S.D., and Brandeis, L.D. 1890. The right to privacy. Harvard Law Review, 4(5): 193-220.

[68] Westin, A.F. 2003. Social and political dimensions of privacy. Journal of Social Issues, 59(2), 431-453.

[69] Woodley, C. and Silvestri, M. 2014. The Internet is forever: student indiscretions reveal the need for effective social media policies in academia. *American Journal of Distance Education*, 28, 2, (2014), 126-138. DOI: 10.1080/08923647.2014.896587

[70] Zhang, Y. and Wildemuth, B.M. 2009 Qualitative analysis of content. In *Applications of social research methods to questions in Information and Library Science,* B. Wildemuth Ed., Libraries Unlimited, Westport, CT, 308-319.

[71] Zuckerman, E. 2014, October 6. Helen Nissenbaum on Ad Nauseum, resistance through obfuscation, and weapons of the weak. [Web log post]. Retrieved from http://www.ethanzuckerman.com/blog/2014/10/06/helen-nissenbaum-on-ad-nauseum-resistance-through-obfuscation-and-weapons-of-the-weak/